

SIPSO: Selectively Informed Particle Swarm Optimization Based on Mutual Information to Determine SNP-SNP Interactions

Wenxiang Zhang¹, Junliang Shang^{1,2}(✉), Huiyu Li¹, Yingxia Sun¹,
and Jin-Xing Liu¹

¹ School of Information Science and Engineering,
Qufu Normal University, Rizhao 276826, China

{zhangwenxiang94, shangjunliang110, huizilovereasing,
sunyingxia026}@163.com, sdcavell@126.com

² Institute of Network Computing,
Qufu Normal University, Rizhao 276826, China

Abstract. Interactive effects of Single Nucleotide Polymorphisms (SNPs), namely, SNP-SNP interactions, have been receiving increasing attention in understanding the mechanism underlying susceptibility to complex diseases. Though many works have been done for their detection, the algorithmic development is still ongoing due to their computational complexities. In this study, we apply selectively informed particle swarm optimization (SIPSO) to determine SNP-SNP interactions with mutual information as its fitness function. The highlights of SIPSO are the introductions of scale-free networks as its population structure, and different learning strategies as its interaction modes, considering the heterogeneity of particles. Experiments are performed on both simulation and real data sets, which show that SIPSO is promising in inferring SNP-SNP interactions, and might be an alternative to existing methods. The software package is available online at <http://www.bdmweb.cn/index.php?m=content&c=index&a=show&catid=37&id=99>.

Keywords: SNP-SNP interactions · Particle swarm optimization · Selectively informed · Mutual information · Scale-free network

1 Introduction

SNP-SNP interactions have been receiving increasing attention in understanding the mechanism underlying susceptibility to complex diseases, such as heart disease, Alzheimer's disease, cancer, type 2 diabetes and many others [1]. Detection of SNP-SNP interactions is therefore becomes an urgent, front-burner problem in the field of Bioinformatics. Though lots of algorithms have been proposed for inferring SNP-SNP interactions, the algorithmic development is still ongoing due to their computational and mathematical complexities, including the fitness function that decides how well an SNP combination contributes to the phenotype, the limitation of prior knowledge in interpreting complex genetic architecture of a disease, the intensive computational burden imposed by the enormous search space resulted from the problems of "high

dimensional and small sample size” and “combinatorial explosion”, which obviously hinder further applications of current algorithms in genome-wide association studies.

Recently, several particle swarm optimization (PSO) based methods have been proposed for determining SNP-SNP interactions [2–13]. Yang *et al.* [2] used a binary PSO to evaluate the risk of breast cancer, with odds ratio as its fitness function, which is known as OR-BPSO for short. Based on the OR-BPSO, Chang *et al.* [3] presented a odds ratio discrete binary PSO (OR-DBPSO) for inferring SNP-SNP interactions with the quantitative phenotype. Chuang *et al.* [4] also combined PSO and odds ratio to explore combinations of SNP-SNP interactions in breast cancer. They then developed a chaotic PSO (CPSO) for breast cancer association studies and indeed identified a SNP combination [5]. In order to enhance the reliability of the PSO in determining SNP-SNP interactions associated with breast cancer, they improved the PSO (IPSO) and proved that the IPSO is highly reliable than the OR-BPSO [6]. Recently, they presented a gauss chaotic map PSO (Gauss-PSO) and used it to detect the best association with breast cancer [7]. They confirmed that the Gauss-PSO was capable of identifying higher difference values between cases and controls than both the PSO and the CPSO. By analyzing the performance of the PSO and the CPSO, Yang *et al.* [8] presented a double-bottom chaotic map PSO (DBM-PSO) to overcome their respective limitations. Later, the DBM-PSO is used to infer SNP-SNP interactions based on *chi*-square test [9]. Hwang *et al.* [10] developed a complementary-logic PSO (CLPSO) to increase the efficiency of determining significant SNP-SNP interactions in case-control study. Wu *et al.* [11] proposed a PSO based algorithm and found SNP interactions of rennin-angiotensin system genes against hypertension. Nevertheless, above mentioned methods mainly focus on identifying a best genotype-genotype of a SNP-SNP interaction among possible genotype combinations of SNP combinations, but not several SNP-SNP interactions among possible SNP combinations, which may lose important clues for the exploration of causative factors of complex diseases. In addition, limited sample size of SNP data obviously affects computational accuracies of these methods. More recently, Ma *et al.* [12] proposed PSOMiner for the detection of SNP-SNP interactions, which is a generalized PSO algorithm based on the fitness function of mutual information. Shang *et al.* [13] presented an improved method IOBLPSO for detecting SNP-SNP interactions by using opposition-based learning PSO. IOBLPSO has ensured the ability of global searching and prevented premature convergence. However, above methods treated all particles equally and adopt a single learning strategy, overlooking the heterogeneity of particles. Gao *et al.* [14] employed scale-free networks to represent the population structure of swarms and proposed a selectively informed PSO (SIPSO), where particles select different learning strategies based on their degrees: a densely connected hub particle gets full information from all of its neighbors (fully-informed) while a non-hub particle with low degree can only follow a single yet best performed neighbor (single-informed). Experiment results show that SIPSO is able to significantly improve the optimization performance.

In this paper, we combine SIPSO and mutual information to determine SNP-SNP interactions. The highlights of SIPSO are the introductions of scale-free networks as its population structure, two different learning strategies as its interaction modes, and mutual information as its fitness function. Experiments are performed on several real

and simulation data sets. Results demonstrate that SIPSO is promising in inferring SNP-SNP interactions, and might be an alternative to existing methods.

2 Methods

2.1 Scale-Free Network

A scale-free network is a network whose degree distribution follows a power law, at least asymptotically. So far, many networks have been reported to be scale-free, for instance, world wide web, internet, neural networks, and so on. Liu et al. studied the influence of scale-free population structure on the performance of PSO, showing that the scale-free PSO outperforms the traditional PSO, due to only a few particles being densely connected hubs and most particles being low degree non-hubs, resulting in high heterogeneity of degrees of particles [14, 15].

In SIPSO, particles are considered as nodes of a scale-free network. This network is generated by Barabasi-Albert model [16], which incorporates two ingredients including growth and preferential attachment. Specifically, at first m_0 fully connected particles are constructed as a complete network; at each time step a new particle is added to the network, which is connected to m ($m < m_0$) nodes with a specified probability. This probability p_i of the new particle being connected to an existing particle i depends upon degree of i , i.e., k_i , which can be written as

$$p_i = \frac{k_i}{\sum_j k_j}, \quad (1)$$

where j runs over all the existing particles. Here m_0 is set to 4 and m is set to 2.

2.2 Mutual Information

The SIPSO introduces mutual information to decide which SNP combinations are SNP-SNP interactions, and to measure how much the effects of captured SNP-SNP interactions to the phenotype are, since mutual information has widely been used as a promising measure for measuring dependence of two variables without a complex modeling. The formula of mutual information can be written as

$$MI(S; Y) = H(S) + H(Y) - H(S, Y), \quad (2)$$

where $H(S)$ is the entropy of S , $H(Y)$ is the entropy of Y , and $H(S, Y)$ is the joint entropy of both S and Y , S is a position of a particle indicating a SNP combination, Y is the phenotype.

The entropy and the joint entropy are defined as,

$$H(S) = - \sum_{j_1=1}^3 \cdots \sum_{j_K=1}^3 (p(s_{j_1}, \dots, s_{j_K}) \cdot \log p(s_{j_1}, \dots, s_{j_K})), \quad (3)$$

$$H(Y) = - \sum_{j=0}^1 (p(y_j) \cdot \log p(y_j)), \quad (4)$$

$$H(S, Y) = - \sum_{j_1=1}^3 \cdots \sum_{j_K=1}^3 \sum_{j=0}^1 (p(s_{j_1}, \dots, s_{j_K}, y_j) \cdot \log p(s_{j_1}, \dots, s_{j_K}, y_j)), \quad (5)$$

where K is the considered order of SNP-SNP interactions, i.e., the number of SNPs in a SNP combination. s is the genotype of a SNP coded as $\{1, 2, 3\}$, corresponding to homozygous common genotype, heterozygous genotype, and homozygous minor genotype. y is the label of a sample coded as $\{0, 1\}$, corresponding to control and case. $p(\cdot)$ is the probability distribution function.

Obviously, higher mutual information value indicates stronger association between the phenotype and the SNP combination.

2.3 Selectively Informed Particle Swarm Optimization (SIPSO)

The SIPSO, first proposed by Gao *et al.* [14] that takes into account both the scale-free population structure and selectively informed learning strategies, is a typical swarm intelligence algorithm that derives the inspiration from the self-organization and adaptation in flocking phenomena. Experimental comparisons showed that SIPSO outperforms several peer algorithms in terms of solution quality, convergence speed, and success rate. We combine SIPSO and mutual information to determine SNP-SNP interactions in this paper.

For SIPSO, let $X_i^t = [x_{i1}^t, \dots, x_{iK}^t]$ and $V_i^t = [v_{i1}^t, \dots, v_{iK}^t]$ be the position and the velocity of particle i at iteration t , where $t \in [1, T]$, $i \in [1, I]$, T is the number of iterations, I is the number of particles, x_{iK}^t is the selected K_{th} SNP index of particle i at iteration t , and $x_{iK}^t \in [1, M]$, M is the number of SNPs in the data set. v_{iK}^t is the velocity of x_{iK}^t . At each iteration, each particle update its position and velocity according to the following equations:

$$\tilde{V}_i^{t+1} = \begin{cases} \eta \cdot \left(V_i^t + \frac{1}{|N(i)|} \sum_{j \in N(i)} \left(U(0, c_1 + c_2) \cdot (P_j^t - X_i^t) \right) \right) & |N(i)| > \lambda \\ \eta \cdot (V_i^t + U(0, c_1) \cdot (P_i^t - X_i^t) + U(0, c_2) \cdot (G_i^t - X_i^t)) & |N(i)| \leq \lambda \end{cases}, \quad (6)$$

$$V_i^{t+1} = \begin{cases} \tilde{V}_i^{t+1} & \tilde{V}_i^{t+1} \in [1 - M, M - 1] \\ U(1 - M, M - 1) \tilde{V}_i^{t+1} & \tilde{V}_i^{t+1} \notin [1 - M, M - 1] \end{cases}, \quad (7)$$

$$\tilde{X}_i^{t+1} = X_i^t + V_i^t, \quad (8)$$

$$X_i^{t+1} = \begin{cases} \text{int}(\tilde{X}_i^{t+1}) & \tilde{X}_i^{t+1} \in [1, M] \\ \text{int}(U(1, M)) & \tilde{X}_i^{t+1} \notin [1, M] \end{cases}, \quad (9)$$

$$\eta = \frac{2}{\left| 2 - (c_1 + c_2) - \sqrt{(c_1 + c_2)^2 - 4 \cdot (c_1 + c_2)} \right|}. \quad (10)$$

Here, P_i^t is the best historical position of particle i until iteration t , G_i^t is the best historical position of neighbors of particle i at iteration t . c_1 and c_2 are the acceleration coefficients controlling how far a particle moves at an iteration. η is the learning rate. $U(a, b)$ is a random number drawn at each iteration from the uniform distribution $[a, b]$. $N(i)$ is the neighbors of particle i and $|N(i)|$ is the degree of particle i (e.g. the number of neighbors). $\text{int}(\cdot)$ is an integral function. λ is the threshold to determine a particle fully-informed or single-informed.

For constructing a positive feedback mechanism, P_i^t and G_i^t should be updated at every iteration. Specifically, whether P_i^{t+1} is updated to X_i^t or P_i^t depends on their mutual information values, which can be defined as,

$$P_i^{t+1} = \begin{cases} P_i^t & MI(P_i^t; Y) = \max(MI(P_i^t; Y), MI(X_i^t; Y)) \\ X_i^t & MI(X_i^t; Y) = \max(MI(P_i^t; Y), MI(X_i^t; Y)) \end{cases}. \quad (11)$$

Similarly, by evaluating mutual information values of $N(i)$ and G_i^t , G_i^{t+1} is updated to one of them with the higher value, which can be written as,

$$G_i^{t+1} = \begin{cases} G_i^t & MI(G_i^t; Y) = \max(MI(G_i^t; Y), MI(N(i); Y)) \\ N(i) & MI(N(i); Y) = \max(MI(G_i^t; Y), MI(N(i); Y)) \end{cases}. \quad (12)$$

While completing the iteration process, P_i^T with descending mutual information values are recorded as the detected SNP-SNP interactions of SIPSO.

3 Results and Discussion

3.1 Simulation and Real Data Sets

We exemplify 4 benchmark models of SNP-SNP interactions for the experiments [17–21], detailed parameters of which are recorded in Table 1. Specifically, Model 1 is a model that display both marginal effects and interactive effect, the penetrance of which increases only when both SNPs have at least one minor allele [17, 18]; Model 2 is a model showing both marginal effects and interactive effect, the additional minor allele at each locus of which does not further increase the penetrance [17]. Model 3 is also a model displaying both marginal effects and interactive effect, which assumes that the minor allele in one SNP has the marginal effect, however, the marginal effect is inversed while minor alleles in both SNPs are present [17]; Model 4 is a model that shows only interactive effects, which is directly cited from the reference [20].

Model 4 is exemplified here since it provides a high degree of complexity to challenge ability of a method in detecting SNP-SNP interactions. For each model, 50 data sets are simulated by *epiSIM* [22], each containing 4000 samples with the ratio of cases and controls being 1. For each data set, random SNPs are set with their MAFs chosen from $[0.05, 0.5]$ uniformly.

A real data set of age-related macular degeneration (AMD) is used for testing the practical ability of SIPSO. AMD, refers to pathological changes in the central area of the retina, is the most important cause of irreversible visual loss in elderly populations, and is considered as a complex disease having multiple SNP-SNP interactions [18]. The AMD data set contains 103611 SNPs genotyped with 96 cases and 50 controls, which has been widely used as a benchmark data set in the field of testing methods of detecting SNP-SNP interactions [12, 13, 17, 18, 21, 23–25].

Table 1. Details of 4 models of SNP-SNP interactions. Prevalence is the proportion of samples that occur a disease. Penetrance is the probability of the occurrence of a disease given a particular genotype. $MAF(\alpha)$ is the minor allele frequency of α . AA , Aa , and aa are homozygous common genotype, heterozygous genotype, and homozygous minor genotype.

Model	MAF(a)	MAF(b)	Prevalence	Penetrance function			
				Genotypes (SNP A)	Genotypes (SNP B)		
					BB	Bb	bb
Model 1	0.300	0.200	0.100	AA	0.087	0.087	0.087
				Aa	0.087	0.146	0.190
				aa	0.087	0.190	0.247
Model 2	0.400	0.400	0.050	AA	0.042	0.042	0.042
				Aa	0.042	0.061	0.061
				aa	0.042	0.061	0.061
Model 3	0.400	0.200	0.010	AA	0.009	0.009	0.009
				Aa	0.013	0.006	0.006
				aa	0.013	0.006	0.006
Model 4	0.400	0.400	0.171	AA	0.068	0.299	0.017
				Aa	0.289	0.044	0.285
				aa	0.048	0.262	0.174

3.2 Experiments on Simulation Data Sets

Three PSO based methods for inferring SNP-SNP interactions are used for this comparison study with the SIPSO. They are DBM-PSO, PSOMiner and IOBLPSO. For a fair comparison, parameter settings of these methods are the same. Specifically, the number of particles I and the number of iterations T are respective set to 200 and 20; acceleration factors c_1 and c_2 are chosen their appropriate settings 2.05 based on previous extensive analysis [26]; the inertia weights for compared methods are set to 0.65; the threshold to determine a particle fully-informed or single-informed λ is set to 4.

Detection power is used to evaluate the performance of SIPSO by applying SIPSO on four groups of simulation data sets, each of which consists of a SNP-SNP interaction model. Power of these compared methods on simulation data sets is reported in Fig. 1.

It is seen that performance of the SIPSO is comparable, and sometimes outperforms to those of compared methods. Specifically, power of the SIPSO is much higher than that of DBM-PSO and PSOMiner; nevertheless, its power is behind that of IOBLPSO in three models except the Model 3, since the IOBLPSO introduces the strategy of opposition-learning, which implies that introducing opposition-learning to SIPSO might also improve the performance of SIPSO, which is a direction; the DBM-PSO has the worst performance on all models because it only focus on identifying genotype combinations among SNP combinations; the PSOMiner is indeed a general PSO that used in determining SNP-SNP interactions, and hence has poor power.

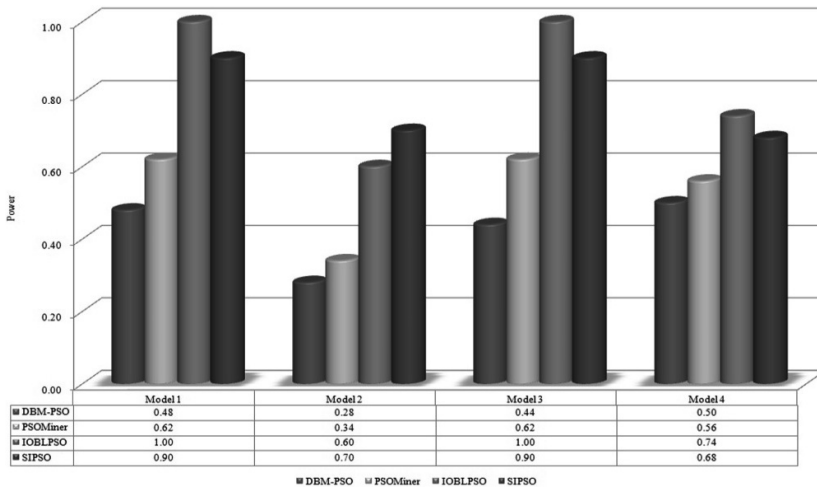


Fig. 1. Power of compared methods on simulation data sets.

3.3 Application to Real Data Set

The SIPSO is applied on AMD data set four times with parameter settings (I, T) being $(10000, 500)$, $(10000, 1000)$, $(20000, 500)$ and $(20000, 1000)$. The top 10 captured SNP-SNP interactions that might be associated with the AMD are reported in Table 2 with descending mutual information values.

It is interesting that all reported SNP-SNP interactions contain either rs380390 or rs1329428. This is because that these two SNPs have strongest main effects among all tested SNPs, and have already proofed to be significantly associated with AMD [18]. Both of them are in the *CFH* gene, and there are biologically plausible mechanisms for the involvement of *CFH* in AMD. *CFH* is a regulator that activates the alternative pathway of the complement cascade, the mutations in which can lead to an imbalance in normal homeostasis of the complement system [24]. This phenomenon is thought to account for substantial tissue damage in AMD [23, 25, 27].

Other SNPs listed in the second column of the table might be identified due to their partners, i.e., rs380390 and rs1329428, since strong main effects of them leads to their combinations with other SNPs almost displaying strong interactive effect. This also

Table 2. Top 10 captured SNP-SNP interactions associated with AMD. *CFH*: complement factor H. *MPP7*: palmitoylated membrane protein 7. *PEBP4*: phosphatidylethanolamine binding protein 4. *R3HDM1*: R3H domain (binds single-stranded nucleic acids). N/A: no gene is available. Chr: Chromosome.

SNP 1				SNP 2				Mutual information value
Index	Name	Gene	Chr	Index	Name	Gene	Chr	
43748	rs380390	<i>CFH</i>	1	80178	rs1363688	N/A	5	0.2966
54108	rs1329428	<i>CFH</i>	1	30550	rs10489076	N/A	4	0.2650
43748	rs380390	<i>CFH</i>	1	51222	rs10508731	<i>MPP7</i>	10	0.2647
43748	rs380390	<i>CFH</i>	1	8764	rs1345488	N/A	18	0.2640
54108	rs1329428	<i>CFH</i>	1	51987	rs2466215	<i>PEBP4</i>	8	0.2636
43748	rs380390	<i>CFH</i>	1	49642	rs10521121	N/A	17	0.2608
43748	rs380390	<i>CFH</i>	1	54213	rs2829015	N/A	21	0.2593
54108	rs1329428	<i>CFH</i>	1	57441	rs1917173	N/A	10	0.2470
43748	rs380390	<i>CFH</i>	1	34447	rs961360	<i>R3HDM1</i>	2	0.2468
54108	rs1329428	<i>CFH</i>	1	64573	rs717246	N/A	1	0.2435

indicates that SIPSO is sensitive to those SNPs displaying strong main effects. Further studies with the use of large scale case-control samples are needed to confirm whether these SNPs have true associations with AMD. We hope that several clues could be provided from this paper for the exploration of causative factors of AMD.

4 Conclusions

SNP-SNP interactions, also known as epistatic interactions or epistasis, is nonlinear interactive effects of SNPs, which is now believed to be one of the causative patterns of complex diseases, and is recognized fundamentally important for understanding the mechanism of disease causing genetic variation. Though many works have been done for inferring SNP-SNP interactions, the algorithmic development is still ongoing due to their mathematical and computational challenges. For instance, the limited fitness functions, the complex genetic architecture, and the intensive computational burden. By considering the advantages of the PSO including simplicity, effectiveness and low computational cost, we introduce the SIPSO with the mutual information to determine SNP-SNP interactions in this manuscript. The SIPSO improves the traditional PSO by modifying the population structure as a scale-free network and altering the interaction modes to different learning strategies. That is to say, learning strategy of each particle depends on its degree: a densely connected hub particle gets full information from all of its neighbors while a non-hub particle with low degree can only follow a single yet best performed neighbor. Experiments are performed on both simulation and real data sets, which show that SIPSO is promising in inferring SNP-SNP interactions, and might be an alternative to existing methods. The Matlab version of SIPSO software package is available online at <http://www.bdmb-web.cn/index.php?m=content&c=index&a=show&catid=37&id=99>.

There are several merits and demerits for SIPLO introducing to identify SNP-SNP interactions. The first advantage is that the SIPSO taking into account the individual's heterogeneity, could balance the exploration and the exploitation in the optimization process. The second advantage is that mutual information is an effective and efficient fitness function in measuring SNP-SNP interactions. Nevertheless, it still has lots of limitations. Firstly, it is only a beneficial exploration of SIPSO applying to determine SNP-SNP interactions, its performance is simply comparable to compared methods, and sometime inferior to compared methods, which indicates that the SIPSO needs more improvement while applying to the field of Bioinformatics. Secondly, SIPSO is sensitive to those SNPs displaying strong main effects. These limitations will inspire us to continue working in the future.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (61502272, 61572284, 61572283), the Scientific Research Reward Foundation for Excellent Young and Middle-age Scientists of Shandong Province (BS2014DX004), the Science and Technology Planning Project of Qufu Normal University (xkj201410), the Opening Laboratory Fund of Qufu Normal University (sk201416), the Scientific Research Foundation of Qufu Normal University (BSQD20130119), The Innovation and Entrepreneurship Training Project for College Students of China (201510446044), The Innovation and Entrepreneurship Training Project for College Students of Qufu Normal University (2015A058, 2015A059).

References

1. Maher, B.: The case of the missing heritability. *Nature* **456**(7218), 18–21 (2008)
2. Yang, C.-H., Chang, H.-W., Cheng, Y.-H., Chuang, L.-Y.: Novel generating protective single nucleotide polymorphism barcode for breast cancer using particle swarm optimization. *Cancer Epidemiol.* **33**(2), 147–154 (2009)
3. Chang, H.-W., Yang, C.-H., Ho, C.-H., Wen, C.-H., Chuang, L.-Y.: Generating SNP barcode to evaluate SNP–SNP interaction of disease by particle swarm optimization. *Comput. Biol. Chem.* **33**(1), 114–119 (2009)
4. Chuang, L.-Y., Lin, M.-C., Chang, H.-W., Yang, C.-H.: Analysis of SNP interaction combinations to determine breast cancer risk with PSO. In: 2011 IEEE 11th International Conference on Bioinformatics and Bioengineering (BIBE), pp. 291–294. IEEE (2011)
5. Chuang, L.-Y., Chang, H.-W., Lin, M.-C., Yang, C.-H.: Chaotic particle swarm optimization for detecting SNP–SNP interactions for CXCL12-related genes in breast cancer prevention. *Eur. J. Cancer Prev.* **21**(4), 336–342 (2012)
6. Chuang, L.-Y., Lin, Y.-D., Chang, H.-W., Yang, C.-H.: An improved PSO algorithm for generating protective SNP barcodes in breast cancer. *PLOS One* **7**(5), e37018 (2012)
7. Chuang, L.-Y., Lin, Y.-D., Chang, H.-W., Yang, C.-H.: SNP-SNP interaction using gauss chaotic map particle swarm optimization to detect susceptibility to breast cancer. In: 2014 47th Hawaii International Conference on System Sciences (HICSS), pp. 2548–2554. IEEE (2014)
8. Yang, C.-H., Tsai, S.-W., Chuang, L.-Y., Yang, C.-H.: An improved particle swarm optimization with double-bottom chaotic maps for numerical optimization. *Appl. Math. Comput.* **219**(1), 260–279 (2012)

9. Yang, C.-H., Lin, Y.-D., Chuang, L.-Y., Chang, H.-W.: Double-bottom chaotic map particle swarm optimization based on chi-square test to determine gene-gene interactions. *BioMed Res. Int.* **2014**, 10 (2014)
10. Hwang, M.-L., Lin, Y.-D., Chuang, L.-Y., Yang, C.-H.: Determination of the SNP-SNP interaction between breast cancer related genes to analyze the disease susceptibility. *Int. J. Mach. Learn. Comput.* **4**(5), 468–473 (2014)
11. Wu, S.-J., Chuang, L.-Y., Lin, Y.-D., Ho, W.-H., Chiang, F.-T., Yang, C.-H., Chang, H.-W.: Particle swarm optimization algorithm for analyzing SNP-SNP interaction of renin-angiotensin system genes against hypertension. *Mol. Biol. Rep.* **40**(7), 4227–4233 (2013)
12. Ma, C., Shang, J., Li, S., Sun, Y.: Detection of SNP-SNP interaction based on the generalized particle swarm optimization algorithm. In: 2014 8th International Conference on Systems Biology (ISB), pp. 151–155. IEEE (2014)
13. Shang, J., Sun, Y., Li, S., Liu, J.-X., Zheng, C.-H., Zhang, J.: An improved opposition-based learning particle swarm optimization for the detection of SNP-SNP interactions. *BioMed Res. Int.* **2015**, 12 (2015)
14. Gao, Y., Du, W., Yan, G.: Selectively-informed particle swarm optimization. *Scientific reports* **5** (2015)
15. Liu, C., Du, W.B., Wang, W.X.: Particle swarm optimization with scale-free interactions. *PLOS One* **9**(5), e97822 (2014)
16. Barabasi, A.L., Albert, R.: Emergence of scaling in random networks. *Science* **286**(5439), 509–512 (1999)
17. Zhang, Y., Liu, J.S.: Bayesian inference of epistatic interactions in case-control studies. *Nat. Genet.* **39**(9), 1167–1173 (2007)
18. Tang, W., Wu, X., Jiang, R., Li, Y.: Epistatic module detection for case-control studies: a Bayesian model with a Gibbs sampling strategy. *PLoS Genet.* **5**(5), e1000464 (2009)
19. Frankel, W.N., Schork, N.J.: Who's afraid of epistasis? *Nat. Genet.* **14**(4), 371–373 (1996)
20. Li, W., Reich, J.: A complete enumeration and classification of two-locus disease models. *Hum. Hered.* **50**(6), 334–349 (2000)
21. Shang, J., Zhang, J., Sun, Y., Liu, D., Ye, D., Yin, Y.: Performance analysis of novel methods for detecting epistasis. *BMC Bioinform.* **12**(1), 475 (2011)
22. Shang, J., Zhang, J., Lei, X., Zhao, W., Dong, Y.: EpiSIM: simulation of multiple epistasis, linkage disequilibrium patterns and haplotype blocks for genome-wide interaction analysis. *Genes Genomics* **35**, 305–316 (2013)
23. Shang, J., Zhang, J., Lei, X., Zhang, Y., Chen, B.: Incorporating heuristic information into ant colony optimization for epistasis detection. *Genes Genomics* **34**(3), 321–327 (2012)
24. Shang, J., Zhang, J., Sun, Y., Zhang, Y.: EpiMiner: a three-stage co-information based method for detecting and visualizing epistatic interactions. *Digit. Signal Process.* **24**, 1–13 (2014)
25. Shang, J., Sun, Y., Fang, Y., Li, S., Liu, J.-X., Zhang, Y.: Hypergraph supervised search for inferring multiple epistatic interactions with different orders. In: Huang, D.-S., Jo, K.-H., Hussain, A. (eds.) ICIC 2015. LNCS, vol. 9226, pp. 623–633. Springer, Heidelberg (2015)
26. Clerc, M., Kennedy, J.: The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Trans. Evol. Comput.* **6**(1), 58–73 (2002)
27. Adams, M.K., Simpson, J.A., Richardson, A.J., Guymer, R.H., Williamson, E., Cantsilieris, S., English, D.R., Aung, K.Z., Makeyeva, G.A., Giles, G.G.: Can genetic associations change with age? CFH and age-related macular degeneration. *Hum. Mol. Genet.* **21**(23), 5229–5236 (2012)